



Multiple Pathways to Homothallism in Closely Related Yeast Lineages in the Basidiomycota

Alexandra Cabrita,^a Márcia David-Palma,^{a*} Patrícia H. Brito,^a Joseph Heitman,^b Marco A. Coelho,^{a*} Paula Gonçalves^a

^aApplied Molecular Biosciences Unit-UCIBIO, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

^bDepartment of Molecular Genetics and Microbiology, Duke University, Duke University Medical Center, Durham, North Carolina, USA

ABSTRACT Sexual reproduction in fungi relies on proteins with well-known functions encoded by the mating type (*MAT*) loci. In the Basidiomycota, *MAT* loci are often bipartite, with the *P/R* locus encoding pheromone precursors and pheromone receptors and the *HD* locus encoding heterodimerizing homeodomain transcription factors (Hd1/Hd2). The interplay between different alleles of these genes within a single species usually generates at least two compatible mating types. However, a minority of species are homothallic, reproducing sexually without an obligate need for a compatible partner. Here, we examine the organization and function of the *MAT* loci of *Cystofilobasidium capitatum*, a species in the order Cystofilobasidiales, which is unusually rich in homothallic species. We determined *MAT* gene content and organization in *C. capitatum* and found that it resembles a mating type of the closely related heterothallic species *Cystofilobasidium ferigula*. To explain the homothallic sexual reproduction observed in *C. capitatum*, we examined HD protein interactions in the two *Cystofilobasidium* species and determined *C. capitatum* *MAT* gene expression both in a natural setting and upon heterologous expression in *Phaffia rhodozyma*, a homothallic species belonging to a clade sister to that of *Cystofilobasidium*. We conclude that the molecular basis for homothallism in *C. capitatum* appears to be distinct from that previously established for *P. rhodozyma*. Unlike in the latter species, homothallism in *C. capitatum* may involve constitutive activation or dispensability of the pheromone receptor and the functional replacement of the usual Hd1/Hd2 heterodimer by an Hd2 homodimer. Overall, our results suggest that homothallism evolved multiple times within the Cystofilobasidiales.

IMPORTANCE Sexual reproduction is important for the biology of eukaryotes because it strongly impacts the dynamics of genetic variation. In fungi, although sexual reproduction is usually associated with the fusion between cells belonging to different individuals (heterothallism), sometimes a single individual is capable of completing the sexual cycle alone (homothallism). Homothallic species are unusually common in a fungal lineage named Cystofilobasidiales. Here, we studied the genetic bases of homothallism in one species in this lineage, *Cystofilobasidium capitatum*, and found it to be different in several aspects from those of another homothallic species, *Phaffia rhodozyma*, belonging to the genus most closely related to *Cystofilobasidium*. Our results strongly suggest that homothallism evolved independently in *Phaffia* and *Cystofilobasidium*, lending support to the idea that transitions between heterothallism and homothallism are not as infrequent as previously thought. Our work also helps to establish the Cystofilobasidiales as a model lineage in which to study these transitions.

KEYWORDS Cystofilobasidiales, *Cystofilobasidium*, evolution of sexual reproduction, homothallism, *MAT* loci, *Phaffia*, sexual reproduction in basidiomycetes

In fungi, as in all eukaryotes, sexual reproduction is widespread, and some of the underlying mechanisms and molecular components are conserved even among distant lineages. The specific molecular pathways involved and the recognition systems

Citation Cabrita A, David-Palma M, Brito PH, Heitman J, Coelho MA, Gonçalves P. 2021. Multiple pathways to homothallism in closely related yeast lineages in the Basidiomycota. mBio 12:e03130-20. <https://doi.org/10.1128/mBio.03130-20>.

Editor B. Gillian Turgeon, Cornell University

Copyright © 2021 Cabrita et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Paula Gonçalves, pmz@fct.unl.pt.

*Present address: Márcia David-Palma and Marco A. Coelho, Department of Molecular Genetics and Microbiology, Duke University, Duke University Medical Center, Durham, North Carolina, USA.

This article is a direct contribution from Joseph Heitman, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Brenda Wingfield, University of Pretoria; Tim James, University of Michigan—Ann Arbor; and Paul Dyer, University of Nottingham.

Received 21 December 2020

Accepted 5 January 2021

Published 16 February 2021

responsible for triggering sexual reproduction may nonetheless vary greatly (1). Generally, sexual reproduction occurs through mating of two haploid individuals of the same species possessing distinct mating types (2, 3). Mating types are defined by specific regions of the genome, the mating type (*MAT*) loci, which encode proteins responsible for triggering the major pathways leading to sexual development. Distinct mating types differ in the genetic contents of the *MAT* loci. In basidiomycetes, mating type determination relies on two *MAT* loci (named *P/R* and *HD*) that encode two different classes of proteins (4–6). The *HD* locus contains two divergently transcribed genes encoding homeodomain transcription factors, and the *P/R* locus comprises pheromone receptor- and pheromone precursor-encoding genes, *STE3* and *MFA*, respectively (4, 5, 7). In the Ascomycota, only one *MAT* locus encoding transcription factors is required to determine mating type identity. Hence, the participation of receptors and pheromones in the determination of mating type is found only in the Basidiomycota.

The *HD* and *P/R* *MAT* loci in Basidiomycota can be either genetically linked or unlinked in the genome of a given species. If these loci are unlinked, they may segregate independently during meiosis, leading to the generation of up to four distinct mating types among the haploid progeny, which is the hallmark of the tetrapolar breeding system (4, 5, 7). The bipolar breeding system results from linkage of the two *MAT* loci and yields only two mating types in the haploid progeny (4, 5, 7). Bipolar mating also takes place if one of the two (unlinked) *MAT* loci loses its function in determining mating type identity, which has been reported occasionally for the *P/R* locus (4, 5, 7).

Although heterothallism involving the fusion between compatible mating types as a requirement for sexual reproduction is common, in some fungal species, individuals are universally compatible, meaning that they can undergo sexual reproduction with any other individual or even with or by themselves, a pattern termed homothallism (3, 7, 8). In basidiomycetes, few homothallic organisms have been found and even fewer have had the genetic basis of their homothallism elucidated (1, 3, 7, 8). Two relevant examples of basidiomycete yeasts with distinct molecular strategies that result in a homothallic sexual mode of reproduction are the human-pathogenic yeast *Cryptococcus deneoformans* (9, 10) and *Phaffia rhodozyma* (11). *Cryptococcus deneoformans* exhibits a particular form of homothallism named unisexual reproduction, where cells of the same mating type can either fuse or undergo endoreplication of the entire genome, resulting in a diploid nucleus that subsequently develops into hyphae and basidia and gives rise to four haploid spores that are products of a meiotic event. Interestingly, the key genes for heterothallic reproduction, like *MFA*, *STE3*, and even *HD1* (*SXI1*) and *HD2* (*SXI2*), appear to be dispensable for the unisexual reproduction of some *C. deneoformans* strains (10, 12, 13).

More recently, the genetic basis of homothallism was dissected in *Phaffia rhodozyma* in our laboratory (11). This astaxanthin-producing basidiomycetous yeast (14) belongs to the order Cystofilobasidiales, a lineage with an unusually high proportion of homothallic species (15–20) but comprising also heterothallic species and others for which sexual reproduction has not yet been observed. In *P. rhodozyma*, deletion mutants were used to show that the two pairs of *STE3* and *MFA* genes and the single *HD1/HD2* gene pair present in the genome are all required for robust sexual reproduction (11). The two pairs of pheromone and pheromone receptors turned out to have reciprocal compatibility, a single compatible *STE3* and *MFA* pair being sufficient for sexual reproduction. This is what might be expected if the *P/R* loci of two putative mating types were present in the same genome, in accord with the definition of primary homothallism. The only pair of Hd1 and Hd2 proteins encoded in the genome is also essential for sexual development, but the mode of action is not fully understood (11). This is because Hd1 and Hd2 are usually expected to heterodimerize to form functional transcription factors, but proteins encoded in the same *HD* locus do not normally form dimers. This imposes heterozygosity at the *HD* locus, with dimerization occurring only between proteins encoded by different gene pairs as a condition for sexual development. In line with this, the Hd1 and Hd2 proteins of *P. rhodozyma* that are encoded in the same gene pair do not appear to interact strongly, leading to the tentative conclusion that a

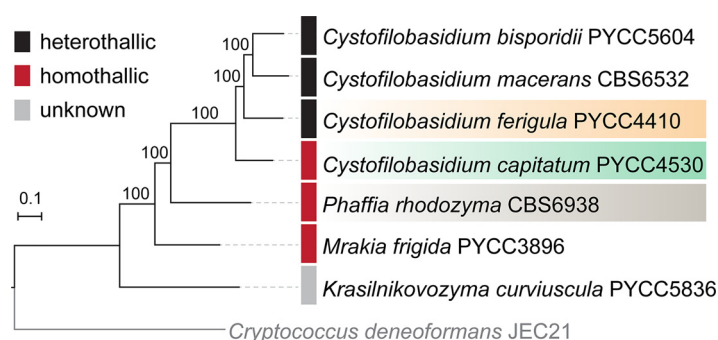


FIG 1 Maximum-likelihood phylogeny reconstructed from the concatenated protein alignments of 1,147 single-copy genes shared across the studied taxa and the outgroup represented by *Cryptococcus deneoformans*. Branch support was assessed by 1,000 replicates of ultrafast bootstrap approximation (UFBoot) with NNI optimization, and branch lengths are given as numbers of substitutions per site.

weak interaction between the two proteins might suffice for function (11). Therefore, the *HD* locus configuration in *P. rhodozyma* is not fully aligned with the concept of primary homothallism, where the presence of two distinct pairs of *HD* genes supporting strong cross dimerization between Hd1 and Hd2 proteins is expected.

Here, we examine in detail the content and function of the *MAT* loci of *Cystofilobasidium capitatum*, a homothallic species closely related to *P. rhodozyma*, in order to understand if there are common features between the molecular bases of homothallism in the two species. We aim to shed some light on the diversity of molecular mechanisms through which homothallism can occur in the phylum Basidiomycota and improve the understanding of the evolution of mating patterns in the *Cystofilobasidiales*.

RESULTS

MAT loci in *Cystofilobasidium capitatum* and *Cystofilobasidium ferigula*.

Cystofilobasidium capitatum and *Cystofilobasidium ferigula* belong to the order *Cystofilobasidiales*, which also contains the genus *Phaffia*, of which two new species were recently described (21), in addition to six other genera (22). The phylogenetic relationships within the order were only recently clarified by a comprehensive genome-based phylogeny (21). This order contains a strikingly large number of homothallic species (approximately one-third of those described so far) (see Table S1 at https://figshare.com/articles/dataset/STables_xlsx/13176422), a sexual mode uncommon in the Basidiomycota (3). In particular, *Phaffia* is composed entirely of homothallic species (21), while *Cystofilobasidium* comprises species with a variety of sexual properties, with *C. capitatum* and *C. intermedium* (23) representing the only two fully homothallic species in the genus. The remaining species of *Cystofilobasidium* are heterothallic, except for *C. macerans*, which comprises strains exhibiting diverse sexual patterns (heterothallic, homothallic, and asexual) and *Cystofilobasidium alribaticum*, for which no sexual reproduction has been observed under the conditions tested (23). The genome-based phylogeny published for the *Cystofilobasidiales* (21) robustly supported *Phaffia* and *Cystofilobasidium* as sister genera, a relationship that is recapitulated in the phylogeny shown in Fig. 1, where a more limited number of species were included.

The availability of draft genomes for several *Cystofilobasidium* species allowed us to examine the *MAT* loci of *C. capitatum* as well as that of the heterothallic species *C. ferigula* (Fig. 2 and Fig. S1) and to compare these with *MAT* loci of homothallic *P. rhodozyma* with regard to gene content and organization (24). In the genome of *C. capitatum* PYCC 4530, a single pair of divergently transcribed homeodomain transcription factors (*HD1* and *HD2*) was found, as well as a single pheromone receptor gene (*STE3*) flanked by two identical pheromone precursor genes (*MFA1a* and *MFA1b*) (Fig. 2). These two sets of genes are located on different scaffolds and encode proteins with

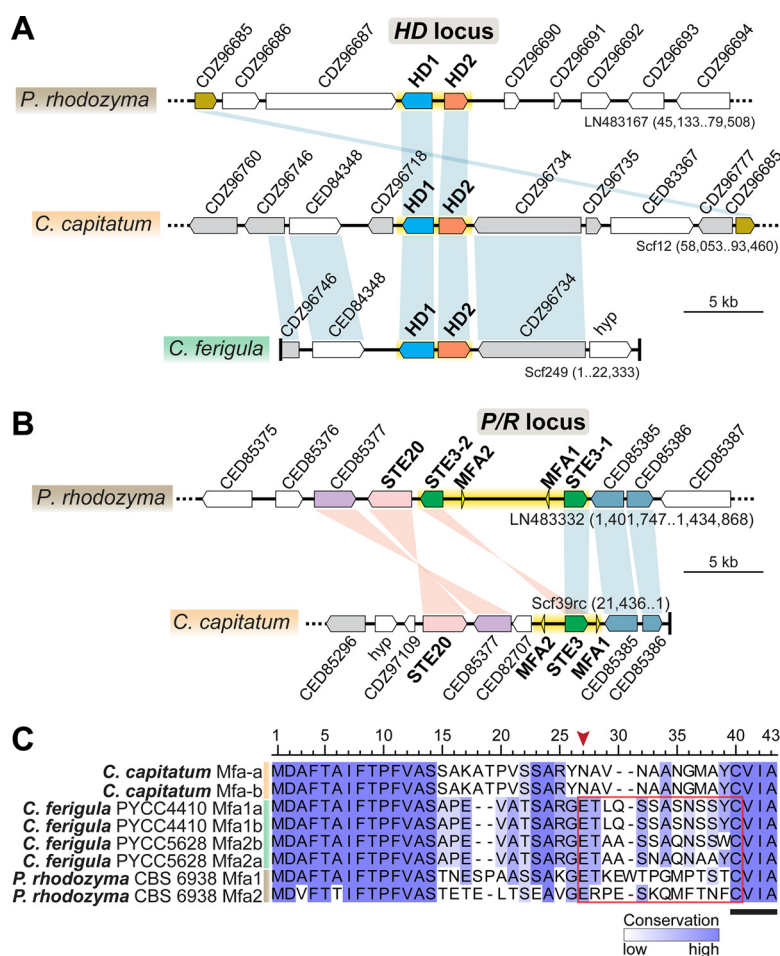


FIG 2 Gene content and organization of the HD (A) and P/R (B) mating type loci in *C. capitatum* and *C. ferigula* (PYCC 4410), compared to those in *P. rhodozyma*. Genes are depicted as arrows denoting the direction of transcription. The genomic regions corresponding to the putative MAT loci are highlighted in yellow. Vertical bars connect orthologs that are in the same (blue) or inverted (pink) orientation. Nonsynthetic genes are shown in white, and genes in gray in *C. capitatum* and *C. ferigula* are found scattered in the corresponding P/R- or HD-containing scaffold in *P. rhodozyma*, suggesting a low level of synteny conservation beyond the core MAT regions. Gene names are based on top BLASTp hits in the *P. rhodozyma* genome, and “hyp” denotes hypothetical proteins. The P/R locus of *C. ferigula* was omitted because the available genome assemblies are too fragmented in these regions to allow for a comparison. (C) Sequence alignment of pheromone precursors, with amino acid positions colored in a blue gradient according to conservation. Sequences proposed as the putative mature pheromones are outlined by a red box, while those resembling the CAAX motif, required for farnesylation, are underlined. In *C. capitatum*, the absence of a conserved position for N-terminal processing (the two charged amino acids indicated by a red arrowhead [25]) precludes the prediction of a mature pheromone sequence.

lengths similar to those of their counterparts in *P. rhodozyma*, and the receptor is predicted to possess seven transmembrane domains, as expected (Fig. 2B and Fig. S2). Both predicted Hd proteins have homeobox domains and nuclear localization signals (NLS) (Fig. S2). However, the MFA genes in *C. capitatum* encode a 41-amino-acid pheromone precursor protein, where no site for N-terminal processing could be identified (Fig. 2C), which may compromise the formation of a mature pheromone (25). Furthermore, analysis of gene synteny conservation between *Phaffia* and *C. capitatum* indicates that the P/R locus of *C. capitatum* may be restricted to the region containing the pheromones/receptors (highlighted in yellow in Fig. 2B), and the HD locus most likely includes only the HD1 and HD2 genes, as observed in many other basidiomycetes (11, 26, 27).

For *C. ferigula*, the genomes of two compatible mating types were obtained (PYCC 4410 and PYCC 5628). Findings concerning MAT gene content were similar to those for

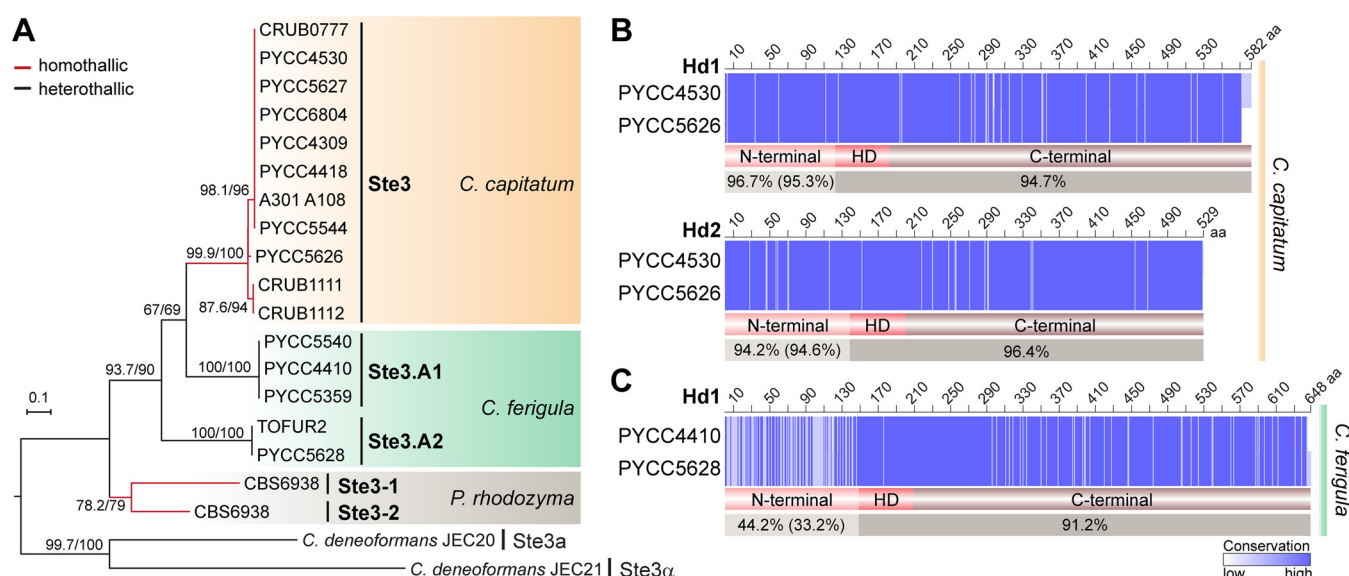


FIG 3 Sequence diversity of *STE3* and *HD* mating type genes in *C. capitatum* and *C. ferigula*. (A) Maximum-likelihood phylogeny of pheromone receptors (*Ste3*) obtained from various strains of *C. capitatum* (homothallic) and *C. ferigula* (heterothallic), along with the previously characterized *Ste3* sequences of *P. rhodozyma* (homothallic) and *C. deneoformans* (heterothallic). The tree was inferred with the LG+G4 model of amino acid substitution and was rooted at the midpoint. Branch support values separated by a slash were assessed by 10,000 replicates of both the Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) and the ultrafast bootstrap approximation (UFBoot). Compared to the other species, the low sequence divergence of *Ste3* sequences among *C. capitatum* strains suggests the absence of functionally distinct, mating type-specific alleles in this species. (B and C) Sequence alignments of the *HD1* and *HD2* gene products of *C. capitatum* and *C. ferigula*. Sequence identity between each pair of *Hd1* and *Hd2* proteins is given for the variable (N termini) and conserved (homeodomain and C termini) regions, with amino acid (aa) positions colored in a blue gradient according to conservation. The comparison between *Hd2* proteins of *C. ferigula* was not performed because the *HD2* gene of PYCC 5628 is fragmented in the current genome assembly. Values in brackets for the N-terminal regions are the average identities as calculated from different allele products (see Fig. S3C for details on the number of alleles). In *C. capitatum*, variable amino acid positions are evenly distributed throughout the length of the *Hd1* and *Hd2* proteins. In contrast, the N-terminal region of *Hd1* in *C. ferigula* is comparatively more variable, as commonly observed in *Hd1* and *Hd2* proteins of other heterothallic basidiomycetes (5, 7, 12, 26).

C. capitatum (Fig. 2 and Fig. S1) and in line with the genetic composition typically found in haploid mating types of basidiomycetes. In *C. ferigula*, two genes encoding pheromone precursors were also found in each of the mating types, but in strain PYCC 5628, the two genes seem to encode slightly different mature pheromones. Because these genes are found in very small contigs in the current genome assemblies, it was not possible to determine their position in relation to that of the *STE3* gene or the exact number of copies in the genome. It is conceivable that additional pheromone genes may become apparent when more-complete assemblies of *C. capitatum* and *C. ferigula* genomes are available. As in *C. capitatum*, the *HD1/HD2* and the *STE3* genes were also found on different scaffolds in both *C. ferigula* strains, but the high level of fragmentation of the current assemblies precludes a precise determination of the length and configuration of the *P/R* and *HD MAT* loci.

While heterothallic species are expected to harbor at least two different mating types with distinct alleles of *MAT* genes, this does not necessarily apply to homothallic species because there is no requirement for an operational self/nonself recognition system in this case. To assess allele diversity of *MAT* genes in the *Cystoflobasidium* species under study, we obtained *MAT* gene sequences for as many strains as possible for both species. For *C. ferigula*, two clearly distinct *STE3* alleles (sharing ~51% amino acid sequence identity) could be recognized among the five strains examined, as expected for a heterothallic species (Fig. 3A). For *C. capitatum*, although several different alleles could be identified in the 11 strains examined, they were much less divergent (sharing ~98% amino acid sequence identity) than *STE3* alleles known to encode proteins with different specificities, like those of *P. rhodozyma* (~50% sequence identity [11]) and *C. ferigula* (Fig. 3A).

The comparison of *C. capitatum* *Hd* proteins suggests the existence of three main *Hd* variants in the species with a degree of divergence between them that is lower

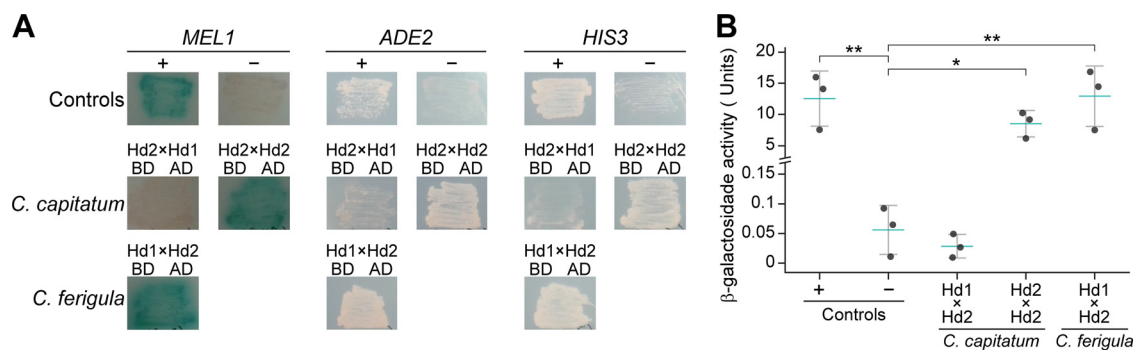


FIG 4 Results of the yeast two-hybrid assay for Hd proteins of *C. capitatum* and *C. ferigula*. (A) Qualitative results from the dimerization of Hd1 and Hd2 of *C. capitatum* and of *C. ferigula*, as well as the homodimerization of Hd2 of *C. capitatum*, through the activation of the reporter genes *MEL1*, *ADE2*, and *HIS3*. The activities of the three reporter genes were tested separately in appropriate media, namely, media containing X- α -Gal and all required supplements for *MEL1* or lacking supplementation with adenine (*ADE2*) or histidine (*HIS3*). The activities of the two latter reporter genes are denoted by the ability of the strains to grow on these media, while *MEL1* expression results in the formation of blue colonies. In *C. ferigula*, the Hd1 protein sequence is derived from strain PYCC 5628 and the Hd2 protein is derived from strain PYCC 4410. (B) Quantitative results of the β -galactosidase assay of interactions shown in panel A. Each data point is the average of the activities measured in two replicate assays; the three data points shown for each strain represent β -galactosidase activities measured in reactions stopped after 2, 6, and 24 h (see also Table S8 at https://figshare.com/articles/dataset/STables_xlsx/13176422). Blue bars denote means and vertical bars standard errors of the means. The interaction between a fusion protein containing the Gal4 activation domain fused to the SV40 large T antigen and a fusion between the Gal4 binding domain and p53 was used as a positive control, while the negative control employed the same Gal4 activation domain fusion in combination with a fusion between the Gal4 binding domain and lamin. Plasmids encoding the positive- and negative-control proteins were provided with the Matchmaker Gold yeast two-hybrid system (TaKaRa Bio, USA). AD, the activation domain of Gal4; BD, the DNA binding domain of Gal4. Significant differences between means were calculated using the Tukey's honestly significant difference (HSD) test (*, $P < 0.05$; **, $P < 0.01$).

than observed for functionally different alleles of heterothallic species (Fig. S3B and S3C). Moreover, for the two proteins that could be examined over their entire lengths, the differences in the sequences are distributed homogeneously throughout the protein (Fig. 3B), as was observed for *P. rhodozyma* (11). This is unlike the divergence observed in Hd1 sequences of *C. ferigula* and of other heterothallic basidiomycetes, which is more extensive and concentrated in the N-terminal region responsible for self/nonself recognition (Fig. 3C).

C. ferigula was reported to be bipolar, as only two different mating types have been identified so far (16). This means that each *STE3* allele is expected to be linked to a single *HD* allele, defining two mating types. However, as shown in Fig. S3C, our analysis uncovered three *HD* alleles, instead of the expected two, and the *HD.B1* allele appears associated with the two receptors in different strains, whereas the *STE3.A1* allele is associated with *HD.B1* and *HD.B3*. From these observations, it seems more likely that *C. ferigula* may have a tetrapolar mating system, in contrast to previous assumptions (16).

Involvement of the Hd proteins of *C. capitatum* in the homothallic sexual cycle.

In heterothallic basidiomycetes, the Hd1 and Hd2 proteins encoded in the *HD* locus control the later stages of sexual reproduction through heterodimerization of nonallelic Hd1 and Hd2 proteins brought together by cell fusion (4, 12, 26). Previous studies in our laboratory concerning the molecular mechanisms of sexual reproduction of *P. rhodozyma* revealed that both the Hd1 and Hd2 proteins are required for normal sporulation (which for this species consists of basidium formation) and likely act through heterodimerization despite the weak nature of their interaction (11).

To understand if *C. capitatum* resembled *P. rhodozyma* in this respect, the yeast two-hybrid assay was employed to assess the ability of the Hd1 and Hd2 proteins of *C. capitatum* to interact with each other. *HD1* and *HD2* cDNAs were isolated from *C. capitatum* strain PYCC 5626 and utilized for the construction of the Gal4 fusion genes for this assay. The results of the assay, presented in Fig. 4, are consistent with a complete absence of interaction between the Hd1 and Hd2 proteins of *C. capitatum*, unlike the results for *P. rhodozyma* (11). Notably, a strong homodimerization was detected for the

TABLE 1 Complementation of *P. rhodozyma* MAT loci mutant strains with *C. capitatum* genes^a

Strain	Sporulation phenotype ^b
<i>P. rhodozyma</i> wild type	+++
<i>HDΔ</i>	–
<i>HDΔ</i> + <i>HD2</i> -PYCC4530	+
<i>HDΔ</i> + <i>HD1</i> / <i>HD2</i> -PYCC4530	+
<i>P/RΔ</i>	–
<i>P/RΔ</i> + <i>STE3</i> / <i>MFA</i> -PYCC4530	–

^aSporulation (basidium formation) patterns were scored qualitatively using the criteria explained below.^bBasidia were counted on the surfaces of 30 colonies per assay; results were scored from two to four assays per strain. –, <5 basidia; +, 5 to 30 basidia; +++, >1,000 basidia.

Hd2 protein from *C. capitatum* (Fig. 4). In *P. rhodozyma*, weak homodimerizations of Hd proteins were also previously observed (11). For *C. ferigula*, a strong heterodimerization of the Hd1 and Hd2 proteins derived from strains of different mating types (PYCC 5628 and PYCC 4410, respectively) was noted (Fig. 4), in line with observations in other heterothallic basidiomycete species (12, 26). Consistently, β -galactosidase activity resulting from activation of the *lacZ* reporter gene was similar to that of the positive control both for heterodimerization of Hd proteins from *C. ferigula* and for homodimerization of Hd2 in *C. capitatum* (Fig. 4B). Homodimerization of Hd1 in *C. capitatum* could not be tested, because the construct of the fusion protein between Hd1 and the Gal4 DNA binding domain could not be stably expressed in the pertinent *Saccharomyces cerevisiae* strain.

The *HD* locus of *C. capitatum* partially complements a *P. rhodozyma* *HD* deletion mutant. The results obtained in the yeast two-hybrid assay suggest that homodimerization of Hd2 may play a role in homothallic sexual development in *C. capitatum*, possibly functionally replacing the usual Hd1/Hd2 heterodimer. To investigate this, we set out to assess the function of the *C. capitatum* *HD* locus by heterologous expression in a *P. rhodozyma* *HDΔ* mutant.

Integration of the complete *HD* locus of *C. capitatum* strain PYCC 4530 in the ribosomal DNA (rDNA) locus of the *P. rhodozyma* *HDΔ* mutant (construct *HDΔ* + *HD1*/*HD2*-PYCC4530) resulted in a very weak but consistent recovery of sporulation (Table 1; Tables S2 and S3 at https://figshare.com/articles/dataset/STables_xlsx/13176422). Because no interaction between the Hd1 and Hd2 proteins could be detected in the yeast two-hybrid assay, but instead strong homodimerization of the Hd2 protein was observed, we subsequently decided to assess whether the expression of *HD2* alone was sufficient to restore the sexual development of the *P. rhodozyma* *HDΔ* mutant. To this end, a construct containing in addition to the intergenic region only a residual (344-bp) 5' portion of the *HD1* gene (excluding the homeodomain; *HDΔ* + *HD2*-PYCC4530) was used to transform the *P. rhodozyma* *HDΔ* mutant. In this mutant, which expresses only the Hd2 protein, sporulation levels like those observed upon transformation of the complete *C. capitatum* *HD* locus were observed (Table 1; Tables S2 and S3 at https://figshare.com/articles/dataset/STables_xlsx/13176422), suggesting that only the Hd2 protein is required for the observed complementation.

The *HD1* and *HD2* genes of *C. capitatum* are differently expressed during growth under sporulation conditions and in *P. rhodozyma* mutants. To substantiate the hypothesis that Hd2 might be the sole important player in the *HD* locus of *C. capitatum*, we compared the expression levels of *HD1* and *HD2* in *C. capitatum* PYCC 5626, under sporulation conditions and in the *P. rhodozyma* mutant containing the complete *HD* locus of *C. capitatum* PYCC 4530 (construct *HDΔ* + *HD1*/*HD2*-PYCC4530). The results, depicted in Fig. 5, show that of the *HD* gene pair, only the *HD2* gene seems to be substantially expressed in the *C. capitatum* strain. Heterologous expression of *HD2* in the *P. rhodozyma* mutant, although much lower than in *C. capitatum*, can easily be detected, while heterologous expression of *HD1* seems to be only vestigial.

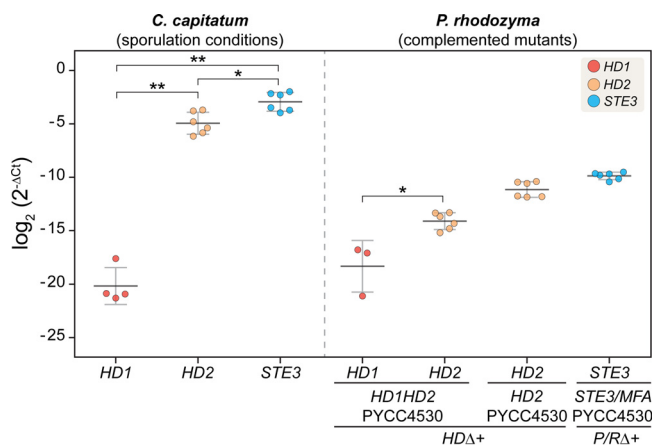


FIG 5 Real-time quantitative PCR results of *MAT* gene expression in *C. capitatum* under sporulation conditions and in complemented *P. rhodozyma* mutants. The expression of *MAT* genes is given as a \log_2 fold difference from the expression of actin in each strain. These results are derived from two biological replicates, each assayed in triplicate, resulting in six data points for each measurement. For *HD1* expression assays, less than six dots are plotted because in some reactions, *HD1* expression was undetectable. Significant differences between the levels of expression of different genes were calculated with the Mann-Whitney test (*, $P < 0.05$; **, $P < 0.01$).

The *P/R* locus of *C. capitatum* does not restore sporulation in a *P. rhodozyma* cognate deletion mutant. Because *Hd* function could be assessed using heterologous expression, a *P. rhodozyma* deletion mutant of both *P/R* clusters (*P/R* Δ mutant) was similarly used as the host for the *P/R* locus of *C. capitatum* strain PYCC 4530, encompassing the *STE3* and *MFA1a* genes and the respective native promoter regions (*P/R* Δ +*STE3/MFA*-PYCC4530). However, integration of the *P/R* locus of *C. capitatum* in the rDNA of the *P. rhodozyma* *P/R* Δ mutant failed to restore sporulation (Table 1; Table S2 at https://figshare.com/articles/dataset/STables_xlsx/13176422), although the *C. capitatum* *STE3* gene was expressed in *P. rhodozyma* (Fig. 5). Interestingly, *STE3* is the most expressed among *MAT* genes of *C. capitatum* PYCC 5626 (Fig. 5), suggesting that it may have a role in sexual reproduction despite our inability to observe it in the heterologous setting.

DISCUSSION

The main aim of this study was to ascertain to what extent common features could be found between the molecular bases of homothallism in different species of the Cystofilobasidiales, a lineage in the Basidiomycota particularly rich in species exhibiting this uncommon sexual behavior. The molecular basis of homothallism was previously dissected in the genetically tractable Cystofilobasidiales species *P. rhodozyma* (11). Here, we characterized the *MAT* loci of a second homothallic species belonging to a sister genus, *C. capitatum*, by examining the structure of the loci in the available genome of strain PYCC 4530 and by comparing *MAT* gene sequences in a total of 11 *C. capitatum* strains. The most striking difference between the *MAT* loci in *P. rhodozyma* and *C. capitatum* is the presence in the latter species of a single pheromone receptor gene, instead of the two distinct and functionally complementary sets of pheromone receptor and pheromone precursor genes found in *P. rhodozyma*. Therefore, the *C. capitatum* *MAT* gene content resembles a haploid mating type of a heterothallic species (4), while that of *P. rhodozyma* is reminiscent of a fusion between two mating types (11). However, as in *P. rhodozyma*, no evidence for functionally distinct variants (alleles) of *MAT* genes that might form different mating types was found in *C. capitatum*, consistent with its homothallic behavior. We subsequently devised various experimental approaches to try to bring to light functional features of *MAT* genes in *C. capitatum*. In these experimental approaches, two *C. capitatum* strains were used; strain PYCC

5626, in addition to sequenced strain PYCC 4530, was employed for the isolation of cDNAs because unlike other strains, it readily sporulates under different experimental conditions.

P. rhodozyma can be described as a primary homothallic basidiomycete, with reciprocal compatibility between the pheromones and receptors of the two *P/R* clusters and with a weak heterodimerization between the only pair of Hd1 and Hd2 proteins being responsible for the triggering of its sexual cycle, with possible hints at homodimerization events (11). Therefore, in *P. rhodozyma*, the *P/R* system seems to work similarly to what might be expected for a heterothallic mating type, while the *HD* locus apparently operates through weak dimerization between adjacently encoded Hd1 and Hd2. This dimerization normally occurs only between proteins encoded by different *HD* alleles. Hence, the weak interaction in *P. rhodozyma* may have evolved to support sexual development in the absence of a second *HD* allele. Because the *MAT* loci in *C. capitatum* have the same gene contents as haploid mating types of heterothallic species, in this case, both the *P/R* and *HD* loci probably underwent changes in their modes of operation to permit homothallism. For the *HD* locus, the complete absence of interaction between the Hd1 and Hd2 proteins, the formation of Hd2 homodimers, complementation results of the *P. rhodozyma* *HDΔ* mutant, and the very low expression of the *HD1* gene suggest that the *HD* locus is relevant for sporulation but also that regulatory functions normally fulfilled by the Hd1/Hd2 heterodimer may have been taken over by the Hd2 homodimer. Hd1 homodimerization was not tested due to technical difficulties. Although it would have been very interesting to be able to test also Hd1 homodimerization, this lost importance when gene expression experiments showed that *HD1* expression was barely detectable. Interestingly, the $\alpha 2$ transcription factor, which is the Hd1 *S. cerevisiae* homolog, is capable of forming both (i) a homodimer that represses the transcription of α -specific genes in haploid α cells and (ii) a heterodimer with the $\alpha 1$ transcription factor (Hd2 homolog), which represses haploid-specific genes in diploid cells (28). This suggests that these proteins can operate both as homodimers and as heterodimers. However, to our knowledge, functional Hd1 or Hd2 homodimers have not been reported so far in basidiomycetes, although there was some evidence for homodimerization of Hd proteins in *P. rhodozyma* (11). Our hypothesis for the mode of action of the *HD* locus predicts that Hd1 might be dispensable. In some heterothallic basidiomycetes, it has been shown that the Hd1 and Hd2 proteins hold different functional domains that are essential for the function of the transcription regulation complex (29–31). One such species is the mushroom *Coprinopsis cinerea*, for which it has been shown that Hd1 contributes to the NLS region, allowing the heterodimer to be transported into the nucleus, while the homeodomain of Hd2 is required for the binding of the complex to DNA (29, 32). Hence, in this case, formation of a heterodimer is required for function. In *C. capitatum*, Hd2 possesses both an NLS and a homeodomain, so in that perspective, it seems possible that Hd2 may indeed play a role in the homothallic life cycle of *C. capitatum* that is sufficient and independent of Hd1. It is possible that this hypothesized function of Hd2 evolved recently in this species, which may explain the fact that *HD1* is not pseudogenized. Alternatively, Hd1 may have acquired a function unrelated to sexual reproduction, which might entail its expression under different conditions and justify its maintenance in the genome. Despite multiple attempts, the methods used to transform *P. rhodozyma* failed to yield transformants of *C. capitatum*, which precluded the possibility of testing the hypothesis regarding the role of Hd2 in this species by deletion of the *HD2* gene.

Introduction of the *P/R* locus into the *P. rhodozyma* *P/RΔ* mutant failed to restore the sporulation of the cognate mutant, even though expression of the *STE3* gene in the heterologous setting could be demonstrated. The explanations for this are presently unclear, since the ability of Ste3-like receptors to activate heterologously a sexual development pathway over a much larger phylogenetic distance was previously demonstrated (33). As to the mode of action of the receptor in its normal setting, the possibility that it is constitutively active should be considered, in line with the fact that the

predicted pheromone precursor genes lack the N-terminal processing signals, likely precluding the formation of an active pheromone. Mutations within the pheromone receptor that can lead to constitutively active receptors have previously been reported, resulting in a bypass of the pheromone for sexual reproduction (25, 33–36). Mutations in the pheromone that allow it to activate the receptor encoded in the same *P/R* cluster are also known (25, 37). In fact, mutations that lead to self-activation or self-compatibility have been proposed to form the basis for the transition from a tetrapolar to a bipolar mating system, or even from a heterothallic to a homothallic mating behavior (3, 5) for some species. If the pheromone system does not operate normally in *C. capitatum*, this may be a reason for the lack of complementation of the *P. rhodozyma* *P/RΔ* mutant by the *C. capitatum* *P/R* locus.

Hence, taken together, our results suggest that *P. rhodozyma* and *C. capitatum* attained homothallic breeding systems through different mechanisms, which is consistent with the hypothesis that the *P. rhodozyma* mating system arose at the origin of the genus, possibly by fusion of *P/R* loci of compatible mating types and the adaptation of the HD dimerization system. *C. capitatum*, on the other hand, has an extant makeup of its *MAT* loci that suggests that it may have evolved from a heterothallic mating type, in which the Hd2 homodimer acquired a prominent role. If a mature pheromone is formed despite the absence of an N-terminal processing signal, it might be that the pheromone-receptor pair in *C. capitatum* has become self-activating. On the other hand, if no mature pheromone is formed, the pheromone receptor gene might have bypassed the pheromone requirement and is now constitutively active. Whatever the case may be, the *P/R* system still likely has a role in some aspects of sexual development, which is in agreement with the fact that the *STE3* gene seems to be highly expressed. Although we cannot discard the hypothesis that the receptor fulfills a role unrelated to sexual reproduction, and while such receptors have been reported (38), they were not associated with pheromone precursor genes.

The different particularities of homothallism in the two Cystofilobasidiales species studied so far are suggestive of remarkable levels of plasticity in the evolution of sexual reproduction in this order. It will be interesting to conduct similar studies of other homothallic species of this order, which would allow us to get more complete insight into the array of mechanisms involved, as well as into the possible genomic rearrangement that may have been involved in the transitions between heterothallic and homothallic species. Having uncovered *P. rhodozyma* as a viable host for heterologous expression opens the possibility of assessing the functionality of other *MAT* proteins from uncharacterized species in this order.

MATERIALS AND METHODS

Strains and culture conditions. *P. rhodozyma*, *C. capitatum*, and *C. ferigula* strains (Table S4 at https://figshare.com/articles/dataset/STables_xlsx/13176422) were routinely grown in YPD medium (2% peptone, 1% yeast extract, 2% glucose, 2% agar) at 17 to 20°C. For the preparation of electrocompetent cells, *P. rhodozyma* strains were grown in YPD liquid medium at 20°C, and transformants were incubated at 17°C in selective medium, consisting of YPD plates supplemented with the appropriate antifungal drugs (100 µg/ml of Geneticin and/or 100 µg/ml of zeocin and/or 100 µg/ml of hygromycin B).

Escherichia coli strain DH5α (Gibco-BRL, Carlsbad, CA, USA) was used as a cloning host and was grown at 37°C in LB medium (1% NaCl, 1% tryptone, 0.5% yeast extract, and 2% agar for solid medium) supplemented with ampicillin at 100 µg/ml when appropriate.

DNA extraction, genome sequencing, and assembly. Genomic DNA of *C. ferigula* PYCC 5628 was extracted from single-cell-derived cultures using the ZR fungal/bacterial DNA miniprep kit (ZYMO Research). DNA samples were quantified using Qubit 2.0. Genome sequencing was carried out by commercial providers at the Genomics Unit of the Instituto Gulbenkian de Ciência and at the Sequencing and Genomic Technologies Core Facility of the Duke Center for Genomic and Computational Biology. Two short insert-sized libraries (~500 bp) were prepared with the Nextera kit and subsequently sequenced using the Illumina MiSeq and HiSeq2500 systems to generate paired 300- and 151-nucleotide reads, respectively. After adaptor clipping using Trimmomatic (v0.36), the two sets of reads were assembled with SPAdes (v3.11.1) (39) (with the following parameters: “-careful” to reduce the number of mismatches and short indels in the final assembly and with the k-mer sizes 21, 33, 55, 77, 99, and 127, automatically selected based on read length). Genome assembly quality was assessed by QUAST (v5.0.2) (40), and gene models were predicted *ab initio* using Augustus (41) trained on *Cryptococcus*

neoformans. Genome sequencing data were generated, and the final genome assembly statistics are given in Table S5 at https://figshare.com/articles/dataset/STables_xlsx/13176422.

Identification of mating type genes and synteny analyses. Scaffolds containing *MAT* genes, namely, the homeodomain transcription factors (*HD1/HD2*) and the mating pheromones (*MFA*) and receptors (*STE3*), were identified by BLASTP or TBLASTN in the genome assemblies of *C. capitatum* PYCC 4530 (=CBS 7420; BioProject accession no. [PRJNA371774](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA371774)) and *C. ferigula* PYCC 4410 (BioProject accession no. [PRJNA371786](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA371786)) and in the newly obtained assembly of *C. ferigula* PYCC 5628 (BioProject accession no. [PRJNA371793](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA371793)). Well-annotated *P. rhodozyma* *MAT* proteins (11) were used as search queries. The retrieved *MAT* genes were manually reannotated if required and analyzed further: (i) the transmembrane regions in the pheromone receptor protein were predicted by HMMTOP software (42), (ii) the homeodomain regions in Hd1 and Hd2 proteins were predicted by the InterPro server (43) and compared to the previously characterized homeodomain proteins in the Pfam database, and (iii) nuclear localization signals (NLS) and coiled-coil motifs were identified in the complete Hd1 and Hd2 sequences using, respectively, SeqNLS (with a cutoff of 0.8) (44) and Jpred4 (45) (see Fig. S1B and S1C and Fig. S2). Synteny between *MAT* regions of different strains and species was based on bidirectional BLAST analyses of the corresponding predicted proteins. The short pheromone precursor genes in the genomes of *C. ferigula* and *C. capitatum* were identified manually, as they usually fail automatic detection.

Species and *MAT* gene phylogenies. A phylogenetic analysis representing major lineages within the *Cystofilobasidiales* was inferred on a concatenated protein data set of single-copy core genes of four *Cystofilobasidium* species, *Phaffia rhodozyma* CBS 6938, *Mrakia frigida* PYCC 3896, *Krasilnikovozyma curviuscula* PYCC 5836, and the outgroup *Cryptococcus deneoformans* JEC21 (Table S4 at https://figshare.com/articles/dataset/STables_xlsx/13176422). Orthologous clusters were inferred with all-against-all BLASTP (NCBI BLAST-2.2) searches and the Markov cluster algorithm (OrthoMCL v1.4 [46]) with an inflation factor (F) of 1.5 and a minimum pairwise sequence alignment coverage of 50%, implemented in the GET_HOMOLOGUES package (47). Clusters present in single copy in all analyzed genomes were retained, aligned with MAFFT v7.407 using the G-INS-I method and default parameter values (48), trimmed with BMGE v1.12 using the amino acid option (49), and finally concatenated into a single data set. The species phylogeny was inferred with IQ-TREE v1.6.12 (50) using maximum-likelihood (ML) inference under a LG+G4+I+G4 model of sequence evolution. ModelFinder (51) was used to determine the best-fit model according to the Bayesian information criterion (BIC), and branch support was estimated using ultrafast bootstrap approximation (UFBoot) with NNI optimization (52), both implemented in the IQ-TREE package.

To analyze the *MAT* gene content across strains of *C. capitatum* and *C. ferigula*, protein sequences of the *HD1*, *HD2*, and *STE3* genes were retrieved from the genome assemblies and aligned separately. Conserved regions were used to design primers to amplify the corresponding genomic regions across the available strains of each species (Table S6 at https://figshare.com/articles/dataset/STables_xlsx/13176422). These regions include an ~870-bp region of the *STE3* gene and an ~1.5-kb fragment encompassing the 5' end and intergenic regions of the *HD1* and *HD2* genes (Table S6 at https://figshare.com/articles/dataset/STables_xlsx/13176422 and Fig. S3). Genomic DNA was extracted through a standard phenol-chloroform method (53), and the regions of interest were PCR amplified, purified using an Illustra GFX PCR DNA and gel band purification kit (GE Healthcare Life Sciences), and then sequenced by Sanger sequencing at STAB Vida (Portugal). For phylogenetic analysis of *MAT* genes, amino acid or nucleotide sequences were individually aligned with MAFFT v7.310 (48) using the L-INS-i strategy (–local-pair –maxiterate 1,000), and poorly aligned regions were trimmed with trimAl (–gappyout) (54). The resulting alignments were input into IQ-TREE v1.6.5 (50) ML phylogenies using best-fit models automatically determined by ModelFinder (51) (parameter, –m MFP). The exact model employed in tree reconstruction is given in the respective figure legends. Branch support values were obtained from 10,000 replicates of both UFBoot (52) and the nonparametric variant of the Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) (55). In addition, the option “–bnni” was employed to minimize the risk of overestimating branch supports with UFBoot when in the presence of severe model violations. The resulting phylogenies were midpoint rooted and graphically visualized with iTOL v5.5.1 (56).

Yeast two-hybrid assay. To assess the interaction between the Hd1 and Hd2 proteins, the Matchmaker Gold yeast two-hybrid system kit (TaKaRa) was used. In this system, fusion proteins containing the Gal4 DNA binding domain (BD) are expressed from plasmid pGBKT7 in *MAT α* haploid strain Y2HGold, while fusion proteins containing the Gal4 activation domain (AD) are expressed from plasmid pGADT7 in *MAT α* haploid strain Y187. Diploid strains expressing both AD and BD fusion proteins are used to test interactions and are obtained by mating the haploid strains expressing each of the fusion proteins of interest. Hence, coding DNA sequences of the pertinent *HD* genes were cloned into pGADT7 and pGBKT7, so as to yield plasmids expressing the desired fusion proteins.

Synthetic genes were designed using the coding DNA sequences of the *HD1* and *HD2* genes of *C. ferigula* and were synthesized at Eurofins Genomics (Germany). The *HD1* and *HD2* gene sequences of strains PYCC 4410 and PYCC 5628 from *C. ferigula*, respectively, were adapted to the *S. cerevisiae* codon usage (Fig. S4). The synthetic genes were obtained as inserts of pEX-A258 plasmids.

cDNAs of *HD1* and *HD2* from *C. capitatum* were obtained from total RNA isolated from strain PYCC 5626, briefly as follows. Strain PYCC 5626 was cultivated in GSA medium (0.2% glucose, 0.2% Soytone) in 10% of the flask volume for 8 days at 20°C and 90 rpm (Sartorius Certomat incubation shaker), until inspection under the microscope revealed the presence of teliospores (16). RNA extraction was performed using the ZR fungal/bacterial RNA miniprep kit (ZYMO Research), with a single step of in-column DNase I digestion to free the RNA samples of genomic DNA. cDNA was synthesized from total mRNA using Maxima H Minus reverse transcriptase (by Thermo Scientific) and oligo(dT)₂₀ as the primer, and

synthesis of the second DNA strand was performed using specific primers for the complete *HD1* and *HD2* genes (Table S6 at https://figshare.com/articles/dataset/STables_xlsx/13176422). The fragments corresponding to the *HD1* and *HD2* cDNA sequences of strain PYCC 5626 were sequenced by Sanger sequencing at STAB Vida (Portugal). The protein sequences of Hd1 and Hd2 of strain PYCC 5626 were aligned with those of strain PYCC 4530 (Fig. 3B) using the software MUSCLE (implemented in the software Unipro UGENE v1.30.0 [57]), and the level of intraspecific variability was calculated.

The *HD1* and *HD2* complete synthetic genes from *C. ferigula* and cDNA fragments from *C. capitatum* were amplified using primers that contained 40-bp tails at their 5' ends that correspond to the flanking regions of the multiple-cloning sites present in pGADT7 and in pGBKT7 (Table S6 at https://figshare.com/articles/dataset/STables_xlsx/13176422). Plasmid pGADT7 was then linearized at the multiple-cloning site by digestion with ClaI (Thermo Scientific), while pGBKT7 was linearized by digestion with PstI (Thermo Scientific). *S. cerevisiae* strains Y187 and Y2HGOLD were transformed with inserts and linearized vectors using the transformation method described in the Yeastmaker yeast transformation system 2 protocol. Transformants were selected in appropriate media (yeast nitrogen base without amino acids, with appropriate supplements) (Table S7 at https://figshare.com/articles/dataset/STables_xlsx/13176422) at 30°C. Transformations with linearized pGBKT7 and the *C. capitatum* Hd1 insert resulted in numbers of transformants similar to those observed in other transformations. However, unlike in other transformations, all transformants investigated harbored plasmids without the insert. This was observed only for this combination of vector and insert and persisted in different transformation attempts.

Mating of haploid *S. cerevisiae* strains was performed by incubating a single colony from each of the two haploid transformants to be mated in 200 μ l of YPD medium at 30°C for 24 h at 250 rpm. After incubation, the cells were recovered and thoroughly washed with distilled sterile water and plated on appropriate selective media (Table S7 at https://figshare.com/articles/dataset/STables_xlsx/13176422).

In the Matchmaker Gold yeast two-hybrid system, the *MEL1*, *ADE2*, and *HIS3* reporter genes are used for qualitative assessment of interactions using plate assays, while the *LacZ* reporter gene is used only to quantify the strength of the interaction, through quantification of β -galactosidase activity in cell extracts. To test the activation of the *ADE2* or *HIS3* reporter gene, diploid strains were plated on appropriate selective media without adenine or histidine, respectively, while to test the activation of the yeast *MEL1* reporter gene, encoding a secreted α -galactosidase, haploid transformants (to check for autoactivation) (Fig. S5) and diploid derivatives were plated on appropriate selective medium supplemented with X- α -Gal (5-bromo-4-chloro-3-indolyl- α -D-galactopyranoside; TaKaRa Bio) at a final concentration of 40 μ g/ml.

To quantify the activation of the *lacZ* reporter gene, a β -galactosidase activity assay was performed, using *o*-nitrophenyl β -D-galactopyranoside (ONPG) as a substrate. The assay was performed as described in the *Yeast Protocols Handbook* (58). All reactions were performed in triplicate, so that they could be stopped at 3 different points in time (after 2 h, 6 h, and 24 h) by the addition of 0.4 ml of 1 M Na₂CO₃ to each suspension. Raw data concerning these assays is shown in Table S8 at https://figshare.com/articles/dataset/STables_xlsx/13176422.

Construction of recombinant plasmids and gene deletion cassettes. For the construction of *P. rhodozyma* mutants, recombinant plasmids and gene deletion cassettes were constructed, as follows. Primer sequences (Table S6 at https://figshare.com/articles/dataset/STables_xlsx/13176422) were based on available genome sequences of *P. rhodozyma* strain CBS 6938 (NCBI project no. PRJEB6925 [59]) and *C. capitatum* strain PYCC 4530 (BioProject accession no. PRJNA371774). Plasmids used for the constructions were pJET1.2/blunt (Thermo Scientific); pPR2TN, containing a Geneticin resistance cassette (60); and pBS-HYG (61), containing a hygromycin resistance cassette. All fragments used for cloning and deletion cassettes were amplified by PCR using Phusion high-fidelity DNA polymerase (Thermo Scientific), and the amplified products were purified using either the Illustra GFX PCR DNA and gel band purification kit (GE Healthcare) or the GeneJET gel extraction kit (Thermo Scientific). Constructions were performed using standard molecular cloning methods (62) and *E. coli* strain DH5 α as the host.

Transformation of *P. rhodozyma*. *P. rhodozyma* strains (CBS 6938 and mutants) (Table S4 at https://figshare.com/articles/dataset/STables_xlsx/13176422) were transformed by electroporation with the linearized recombinant plasmids or deletion cassettes, as previously described by Visser et al. (63), with reduction of the amount of DNA used to 2 μ g for the transformations of the complemented *P. rhodozyma* mutants. The electroporation conditions (Gene Pulser II electroporation system; Bio-Rad) consisted of an internal resistance of 1,000 Ω , an electric pulse of 0.8 kV, and a capacitance of 25 μ F, resulting in a pulse ranging from 18 to 20 ms (64, 65). The cells recovered subsequently in YPD liquid medium for at least 2.5 h at 17°C before being plated on YPD medium supplemented with the appropriate antifungal drug and incubated at 17°C. The genotypes of the transformants were determined as previously described (11, 66).

***P. rhodozyma* sporulation (basidium formation) assays.** To determine the ability of the *P. rhodozyma* mutants to sporulate, we performed basidium formation assays where the strains were incubated in DWR solid medium with 0.5% of ribitol (0.5% ribitol and 2.5% agar), as previously described (11). Each basidium formation assay was conducted on 3 plates containing DWR plus 0.5% ribitol. On each plate, 10 colonies of each strain to be tested were spotted. Different strains were employed in each assay, as indicated in Table S2 at https://figshare.com/articles/dataset/STables_xlsx/13176422, but in all assays, the *P. rhodozyma* wild type was used as a positive control. Colonies were examined under the microscope using $\times 100$ magnification after 10 and 20 days of incubation at 18°C, and basidium formation patterns were scored qualitatively. The numbers of basidia counted in experiment E8 for the complemented mutants in the study of the *HD* locus of *C. capitatum* after 20 days of incubation are listed in Table S3 at https://figshare.com/articles/dataset/STables_xlsx/13176422. In all cases, the entire colony was submitted to microscopic observation.

Real-time quantitative PCR to assess expression of the *MAT* genes. Total RNA was extracted from a sporulating culture of *C. capitatum* strain PYCC 5626. Sporulation (teliospore formation in the case of this species) was induced by incubation in GSA liquid medium (0.2% glucose, 0.2% Soytone) in 10% of the volume of the flask at 17°C without agitation until microscopic inspection revealed hyphae and teliospores, the latter being thick-walled resting spores from which basidia arise (16). Total RNA extraction was performed using the ZR fungal/bacterial RNA miniprep kit (ZYMO Research). Complemented *P. rhodozyma* mutants were grown in YPD liquid medium to an optical density at 600 nm (OD_{600}) of 1.0. The cultures were then collected and frozen at -80°C for 1 h before we proceeded with total RNA extraction through a standard TRIzol method. In-column DNase I digestion to free the RNA samples of genomic DNA using the RNA Clean & Concentrator kit (ZYMO Research) was used for all samples, and the absence of genomic DNA was verified by PCR. cDNA was synthesized from total mRNA using Maxima H Minus reverse transcriptase (Thermo Scientific) and oligo(dT)₂₀ as the primer. Real-time PCR was performed using the SensiFAST SYBR No-ROX kit (Bioline, London, UK) with 20- μl reaction mixtures in a Rotor-Gene 6000 Corbett apparatus. The reaction parameters consisted of an initial denaturation step at 95°C for 2 min, followed by 40 cycles of 95°C for 5 s, 57°C for 10 s, and 72°C for 20 s. Two biological replicates were performed, with triplicates performed for each reaction (Table S9 at <https://figshare.com/articles/dataset/STables.xlsx/13176422>). Relative expression of the *MAT* genes was calculated using the $2^{-\Delta C_t}$ method, where ΔC_t is the cycle threshold of the test ($C_{t_{\text{test}}}$) minus the cycle threshold of the reference ($C_{t_{\text{reference}}}$), as described by Livak and Schmittgen (67), and expression values were represented as the $\log_2(2^{-\Delta C_t})$. Mann-Whitney tests were performed to determine if the differences in expression of genes within each strain were statistically significant.

Data availability. Nucleotide sequences have been deposited into the NCBI/EMBL (GenBank) database under the following accession numbers: MT561333 to MT561342 (*C. capitatum* *STE3*), MT561330 to MT561332 (*C. ferigula* *STE3*), and MT592882 to MT592890 and MT592891 to MT592894 (respectively, for the *C. capitatum* and *C. ferigula* 5' ends and intergenic regions of *HD1* and *HD2*). Sequencing reads for *C. ferigula* PYCC 5628 (BioProject accession no. [PRJNA371793](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA371793)) are available in the NCBI SRA database. The *C. ferigula* PYCC 5628 draft genome assembly has been deposited into DDBJ/ENA/GenBank under accession number [MVAN000000000](https://www.ncbi.nlm.nih.gov/nuccore/MVAN000000000).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 0.2 MB.

FIG S2, PDF file, 1 MB.

FIG S3, PDF file, 0.9 MB.

FIG S4, TIF file, 1.9 MB.

FIG S5, TIF file, 0.6 MB.

ACKNOWLEDGMENTS

We are very thankful to Alex Idnurm, Brenda Wingfield, Timothy James, and Paul Dyer for their insightful and thorough reviews of the manuscript.

This work was supported by UCIBIO (Unidade de Ciências Biomoleculares Aplicadas), which is financed by Portuguese funds from the Fundação para a Ciência e Tecnologia, Ministério da Ciência, Tecnologia e Ensino Superior (FCT/MCTES; <https://www.fct.pt/>) grant UID/Multi/04378/2019 and FCT/MCTES grants PTDC/BIA-GEN/112799/2009 (to P.G.), SFRH/BPD/79198/2011 (to M.A.C.), and SFRH/BD/81895/2011 (to M.D.-P.). This work also benefited from the support of the INCD computing infrastructure funded by FCT and FEDER under project 01/SAICT/2016 grant 022153. This work was also supported by NIH/NIAID R37 award AI39115-23 and R01 award AI50113-16.

J.H. is codirector and fellow for the CIFAR program Fungal Kingdom: Threats & Opportunities.

REFERENCES

- Lee SC, Ni M, Li W, Shertz C, Heitman J. 2010. The evolution of sex: a perspective from the fungal kingdom. *Microbiol Mol Biol Rev* 74:298–340. <https://doi.org/10.1128/MMBR.00005-10>.
- Stanton BC, Hull CM. 2007. Mating-type locus control of cell identity, p 59–73. In Heitman J, Kronstad JW, Taylor JW, Casselton LA (ed), *Sex in fungi: molecular determination and evolutionary implications*. ASM Press, Washington, DC.
- Lin X, Heitman J. 2007. Mechanisms of homothallism in fungi and transitions between heterothallism and homothallism, p 35–57. In Heitman J, Kronstad JW, Taylor JW, Casselton LA (ed), *Sex in fungi: molecular determination and evolutionary implications*. ASM Press, Washington, DC.
- Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. 2017. Fungal sex: the basidiomycota. *Microbiol Spectr* 5:FUNK-0046–2016. <https://doi.org/10.1128/microbiolspec.FUNK-0046-2016>.
- Fraser JA, Hsueh Y-P, Findley KM, Heitman J. 2007. Evolution of the mating-type locus: the basidiomycetes, p 19–34. In Heitman J, Kronstad JW, Taylor JW, Casselton LA (ed), *Sex in fungi: molecular determination and evolutionary implications*. ASM Press, Washington, DC.
- Raudaskoski M, Kothe E. 2010. Basidiomycete mating type genes and pheromone signaling. *Eukaryot Cell* 9:847–859. <https://doi.org/10.1128/EC.00319-09>.
- Kües U, James TY, Heitman J. 2011. Mating type in basidiomycetes: unipolar,

- bipolar, and tetrapolar patterns of sexuality, p 97–160. In Pöggeler S, Wöstemeyer J (ed), *The mycota, evolution of fungi and fungal-like organisms*, vol XIV. Springer, Berlin, Germany.
8. Wilson AM, Wilken PM, van der Nest MA, Steenkamp ET, Wingfield MJ, Wingfield BD. 2015. Homothallism: an umbrella term for describing diverse sexual behaviours. *IMA Fungus* 6:207–214. <https://doi.org/10.5598/imafungus.2015.06.01.13>.
 9. Lin X, Hull CM, Heitman J. 2005. Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. *Nature* 434:1017–1021. <https://doi.org/10.1038/nature03448>.
 10. Wang L, Lin X. 2011. Mechanisms of unisexual mating in *Cryptococcus neoformans*. *Fungal Genet Biol* 48:651–660. <https://doi.org/10.1016/j.fgb.2011.02.001>.
 11. David-Palma M, Sampaio JP, Gonçalves P. 2016. Genetic dissection of sexual reproduction in a primary homothallic basidiomycete. *PLoS Genet* 12:e1006110. <https://doi.org/10.1371/journal.pgen.1006110>.
 12. Hull CM, Boily M, Heitman J. 2005. Sex-specific homeodomain proteins Sxi1 α and Sxi2 α coordinately regulate sexual development in *Cryptococcus neoformans*. *Eukaryot Cell* 4:526–535. <https://doi.org/10.1128/EC.4.3.526-535.2005>.
 13. Gyawali R, Zhao Y, Lin J, Fan Y, Xu X, Upadhyay S, Lin X. 2017. Pheromone independent unisexual development in *Cryptococcus neoformans*. *PLoS Genet* 13:e1006772. <https://doi.org/10.1371/journal.pgen.1006772>.
 14. Barredo J, García-Estrada C, Kosalkova K, Barreiro C. 2017. Biosynthesis of astaxanthin as a main carotenoid in the heterobasidiomycetous yeast *Xanthophyllomyces dendrorhous*. *J Fungi (Basel)* 3:44. <https://doi.org/10.3390/jof3030044>.
 15. Kurtzman CP, Fell JW, Boekhout T. 2011. *The yeasts, a taxonomic study*, 5th ed. Elsevier, London, United Kingdom.
 16. Sampaio JP. 2011. *Cystoflobasidium* Oberwinkler & Bandoni (1983), p 1423–1432. In Kurtzman CP, Fell JW, Boekhout T (ed), *The yeasts, a taxonomic study*, 5th ed. Elsevier, London, United Kingdom.
 17. Fell JW, Johnson EA, Scorzetti G. 2011. *Xanthophyllomyces* Golubev (1995), p 1595–1599. In Kurtzman CP, Fell JW, Boekhout T (ed), *The yeasts, a taxonomic study*, 5th ed. Elsevier, London, United Kingdom.
 18. Fell JW. 2011. Mrakia Y. Yamada & Komagata (1987), p 1503–1510. In Kurtzman CP, Fell JW, Boekhout T (ed), *The yeasts, a taxonomic study*, 5th ed. Elsevier, London, United Kingdom.
 19. Sampaio JP. 2011. Tausonia Bab'eva (1998), p 1999–2001. In Kurtzman CP, Fell JW, Boekhout T (ed), *The yeasts, a taxonomic study*, 5th ed. Elsevier, London, United Kingdom.
 20. Takashima M, Nakase T. 2011. *Udeniomyces* Nakase & Takematsu (1992), p 2063–2068. In Kurtzman CP, Fell JW, Boekhout T (ed), *The yeasts, a taxonomic study*, 5th ed. Elsevier, London, United Kingdom.
 21. David-Palma M, Libkind D, Brito PH, Silva M, Bellora N, Coelho MA, Heitman J, Gonçalves P, Sampaio JP. 2020. The untapped Australasian diversity of astaxanthin-producing yeasts with biotechnological potential —*Phaffia australis* sp. nov. and *Phaffia tasmanica* sp. nov. *Microorganisms* 8:1651. <https://doi.org/10.3390/microorganisms8111651>.
 22. Kachalkin AV, Turchetti B, Inácio J, Carvalho C, Mašinová T, Pontes A, Röhl O, Glushakova AM, Akulov A, Baldrian P, Begerow D, Buzzini P, Sampaio JP, Yurkov AM. 2019. Rare and undersampled dimorphic basidiomycetes. *Mycol Prog* 18:945–971. <https://doi.org/10.1007/s11557-019-01491-5>.
 23. Pontes A, Röhl O, Carvalho C, Maldonado C, Yurkov AM, Paulo Sampaio J. 2016. *Cystoflobasidium intermedium* sp. nov. and *Cystoflobasidium alribaticum* f.a. sp. nov., isolated from Mediterranean forest soils. *Int J Syst Evol Microbiol* 66:1058–1062. <https://doi.org/10.1099/ijsem.0.000835>.
 24. Bellora N, Moliné M, David-Palma M, Coelho MA, Hittinger CT, Sampaio JP, Gonçalves P, Libkind D. 2016. Comparative genomics provides new insights into the diversity, physiology, and sexuality of the only industrially exploited tremellomycete: *Phaffia rhodozyma*. *BMC Genomics* 17:901. <https://doi.org/10.1186/s12864-016-3244-7>.
 25. Kües U. 2015. From two to many: multiple mating types in Basidiomycetes. *Fungal Biol Rev* 29:126–166. <https://doi.org/10.1016/j.fbr.2015.11.001>.
 26. Kämpfer J, Reichmann M, Romeis T, Bölker M, Kahmann R. 1995. Multiallelic recognition: nonself-dependent dimerization of the bE and bW homeodomain proteins in *Ustilago maydis*. *Cell* 81:73–83. [https://doi.org/10.1016/0092-8674\(95\)90372-0](https://doi.org/10.1016/0092-8674(95)90372-0).
 27. Maia TM, Lopes ST, Almeida J, Rosa LH, Sampaio JP, Gonçalves P, Coelho MA. 2015. Evolution of mating systems in Basidiomycetes and the genetic architecture underlying mating-type determination in the yeast *Leucosporidium scottii*. *Genetics* 201:75–89. <https://doi.org/10.1534/genetics.115.177717>.
 28. Dranginis AM. 1990. Binding of yeast a1 and a2 as a heterodimer to the operator DNA of a haploid-specific gene. *Nature* 347:682–685. <https://doi.org/10.1038/347682a0>.
 29. Asante-Owusu RN, Banham AH, Böhnert HU, Mellor EJC, Casselton LA. 1996. Heterodimerization between two classes of homeodomain proteins in the mushroom *Coprinus cinereus* brings together potential DNA-binding and activation domains. *Gene* 172:25–31. [https://doi.org/10.1016/0378-1119\(96\)00177-1](https://doi.org/10.1016/0378-1119(96)00177-1).
 30. Schlesinger R, Kahmann R, Kämpfer J. 1997. The homeodomains of the heterodimeric bE and bW proteins of *Ustilago maydis* are both critical for function. *Mol Gen Genet* 254:514–519. <https://doi.org/10.1007/pl00008609>.
 31. Van Diepen LTA, Olson A, Ihrmark K, Stenlid J, James TY. 2013. Extensive trans-specific polymorphism at the mating type locus of the root decay fungus *Heterobasidion*. *Mol Biol Evol* 30:2286–2301. <https://doi.org/10.1093/molbev/mst126>.
 32. Kües U, Göttgens B, Stratmann R, Richardson WV, O'Shea SF, Casselton LA. 1994. A chimeric homeodomain protein causes self-compatibility and constitutive sexual development in the mushroom *Coprinus cinereus*. *EMBO J* 13:4054–4059. <https://doi.org/10.1002/j.1460-2075.1994.tb06722.x>.
 33. Hsueh Y-P, Xue C, Heitman J. 2009. A constitutively active GPCR governs morphogenic transitions in *Cryptococcus neoformans*. *EMBO J* 28:1220–1233. <https://doi.org/10.1038/emboj.2009.68>.
 34. Olesnicki NS, Brown AJ, Dowell SJ, Casselton LA. 1999. A constitutively active G-protein-coupled receptor causes mating self-compatibility in the mushroom *Coprinus*. *EMBO J* 18:2756–2763. <https://doi.org/10.1093/emboj/18.10.2756>.
 35. Boone C, Davis NG, Sprague GF. 1993. Mutations that alter the third cytoplasmic loop of the a-factor receptor lead to a constitutive and hypersensitive phenotype. *Proc Natl Acad Sci U S A* 90:9921–9925. <https://doi.org/10.1073/pnas.90.21.9921>.
 36. Fowler TJ, Mitton MF, Vaillancourt LJ, Raper CA. 2001. Changes in mate recognition through alterations of pheromones and receptors in the multisexual mushroom fungus *Schizophyllum commune*. *Genetics* 158:1491–1503.
 37. Riquelme M, Challen MP, Casselton LA, Brown AJ. 2005. The origin of multiple B mating specificities in *Coprinus cinereus*. *Genetics* 170:1105–1119. <https://doi.org/10.1534/genetics.105.040774>.
 38. Diaz-Valderrama JR, Aime MC. 2016. The cacao pathogen *Moniliophthora roreri* (Marasmiaceae) possesses biallelic A and B mating loci but reproduces clonally. *Heredity (Edinb)* 116:491–501. <https://doi.org/10.1038/hdy.2016.5>.
 39. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
 40. Mikheenko A, Pribelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>.
 41. Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntetically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24:637–644. <https://doi.org/10.1093/bioinformatics/btn013>.
 42. Tusnády GE, Simon I. 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17:849–850. <https://doi.org/10.1093/bioinformatics/17.9.849>.
 43. Mitchell A, Chang H, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong S, Bateman A, Punta M, Attwood T, Sigrist C, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale D, Wu C, Orengo C, Sillitoe I, Mi H, Thomas P, Finn R. 2015. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* 43:D213–D221. <https://doi.org/10.1093/nar/gku1243>.
 44. Lin J, Hu J. 2013. SeqNLS: nuclear localization signal prediction based on frequent pattern mining and linear motif scoring. *PLoS One* 8:e76864. <https://doi.org/10.1371/journal.pone.0076864>.
 45. Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 43:W389–W394. <https://doi.org/10.1093/nar/gkv332>.
 46. Li L, Stoeckert CJ, Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189. <https://doi.org/10.1101/gr.1224503>.
 47. Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* 79:7696–7701. <https://doi.org/10.1128/AEM.02411-13>.

48. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
49. Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10:210. <https://doi.org/10.1186/1471-2148-10-210>.
50. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
51. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>.
52. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>.
53. Gonçalves P, Valério E, Correia C, de Almeida J, Sampaio JP. 2011. Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. *PLoS One* 6:e20739. <https://doi.org/10.1371/journal.pone.0020739>.
54. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
55. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>.
56. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 47:W256–W259. <https://doi.org/10.1093/nar/gkz239>.
57. Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, German GOG, Kandrov D, Rasputin K, Syabro M, Tleukenov T, UGENE team. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28:1166–1167. <https://doi.org/10.1093/bioinformatics/bts091>.
58. Clontech. 2009. Yeast protocols handbook. Clontech, Mountain View, CA. <https://www.takara.co.kr/file/manual/pdf/PT3024-1.pdf>.
59. Sharma R, Gassel S, Steiger S, Xia X, Bauer R, Sandmann G, Thines M. 2015. The genome of the basal agaricomycete *Xanthophyllomyces dendrorhous* provides insights into the organization of its acetyl-CoA derived pathways and the evolution of Agaricomycotina. *BMC Genomics* 16:233. <https://doi.org/10.1186/s12864-015-1380-0>.
60. Verdoes JC, Krubasik P, Sandmann G, Van Ooyen AJJ. 1999. Isolation and functional characterisation of a novel type of carotenoid biosynthetic gene from *Xanthophyllomyces dendrorhous*. *Mol Gen Genet* 262:453–461. <https://doi.org/10.1007/s004380051105>.
61. Niklitschek M, Alcaíno J, Barahona S, Sepúlveda D, Lozano C, Carmona M, Marcolleta A, Martínez C, Lodato P, Baeza M, Cifuentes V. 2008. Genomic organization of the structural genes controlling the astaxanthin biosynthesis pathway of *Xanthophyllomyces dendrorhous*. *Biol Res* 41:93–108. <https://doi.org/10.4067/S0716-97602008000100011>.
62. Sambrook J, Russell D. 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
63. Visser H, Sandmann G, Verdoes JC. 2005. Xanthophylls in fungi. Metabolic engineering of the astaxanthin biosynthetic pathway in *Xanthophyllomyces dendrorhous*, p 257–272. In Barredo JL (ed), *Microbial processes and products*. Humana Press, Totowa, NJ.
64. Rodríguez-Sáiz M, Godío RP, Álvarez V, De La Fuente JL, Martín JF, Barredo JL. 2009. The NADP-dependent glutamate dehydrogenase gene from the astaxanthin producer *Xanthophyllomyces dendrorhous*: use of its promoter for controlled gene expression. *Mol Biotechnol* 41:165–172. <https://doi.org/10.1007/s12033-008-9123-y>.
65. Wery J, Verdoes JC, Van Ooyen AJJ. 1998. Efficient transformation of the astaxanthin-producing yeast *Phaffia rhodozyma*. *Biotechnol Tech* 12:399–405. <https://doi.org/10.1023/A:1008834600770>.
66. Lin G, Bultman J, Johnson EA, Fell JW. 2012. Genetic manipulation of *Xanthophyllomyces dendrorhous* and *Phaffia rhodozyma*. *Methods Mol Biol* 898:235–249. https://doi.org/10.1007/978-1-61779-918-1_16.
67. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta C_T$ method. *Methods* 25:402–408. <https://doi.org/10.1006/meth.2001.1262>.